

2.2.1991-446  
6.8

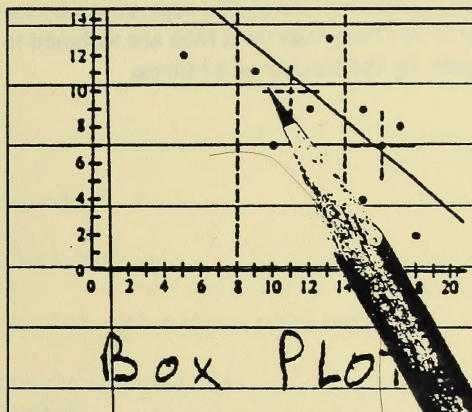
University of Alberta Library



0 1620 3409889 5

# MATHEMATICS 33

Statistics



Unit 6

## Learning Facilitator's Manual



**Distance  
Learning**

**Alberta**  
EDUCATION

## Note

This Mathematics Learning Facilitator's Manual contains answers to teacher-assessed assignments and the final test; therefore, it should be kept secure by the teacher. Students should not have access to these assignments or the final tests until they are assigned in a supervised situation. The answers should be stored securely by the teacher at all times.

Mathematics 33  
Learning Facilitator's Manual  
Unit 6  
Statistics  
Alberta Distance Learning Centre  
ISBN No. 0-7741-0172-5

### ALL RIGHTS RESERVED

Copyright © 1991, the Crown in Right of Alberta, as represented by the Minister of Education, Alberta Education 11160 Jasper Avenue, Edmonton, Alberta, T5K 0L2.

All rights reserved. Additional copies may be obtained from the Learning Resources Distributing Centre.

No part of this courseware may be reproduced in any form including photocopying (unless otherwise indicated), without the written permission of Alberta Education.

Every effort has been made both to provide proper acknowledgement of the original source and to comply with copyright law. If cases are identified where this has not been done, please notify Alberta Education so appropriate corrective action can be taken.



## Topic 1: Bivariate Data

3

1. Circle the correct response.

The science of collecting and classifying data in order to show their significance is called

A. data processing

**B. statistics**

C. technology

D. mathematics

The correct response is B.

5

2. What are the five steps of the statistical process?

The five steps of the statistical process are as follows:

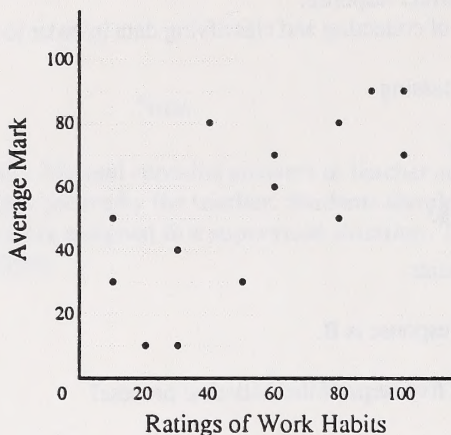
- collection of data
- organizing data in sequential or graphic form such as charts, tables and graphs
- analysis of the organized data
- drawing inferences from the analysis of the organized data
- evaluation for confidence of the inferences

3. The following table of marks shows the relationship between the rating of work habits and the average marks for students at a college.

Students	Rating of Work Habits	Average Mark
Marie	30	10
Sue	10	30
Joe	30	40
Gary	10	50
Gurmeet	50	30
Sandy	60	60
Lou	60	70
Jerry	40	80
Maria	80	80
Rhonda	80	50
Olga	100	70
Joyce	100	90
Chen	90	90
Roy	20	10

5

- a. Draw a scatterplot of the students' average marks against the rating of their work habits. Place the rating of work habits along the horizontal axis.



2

- b. Is there a positive correlation, negative correlation, or no correlation?

There is a strong positive correlation between work habits and the average mark. Students with good work habits generally have higher averages.

②

- c. Discuss the performances of Sandy, Maria, and Chen.

Sandy, Maria, and Chen all have the same ratio of work habit rating to average mark. They are consistent. However, they are not performing equally well. Sandy is much lower than Chen as far as achievement is concerned.

③

- d. How do the results of Gary and Jerry vary?

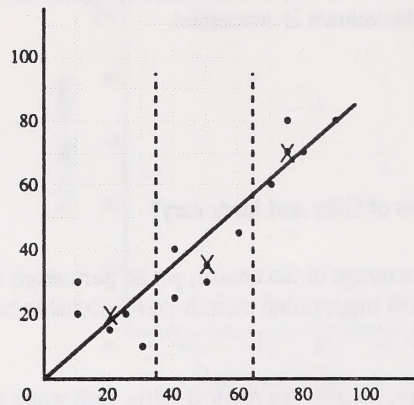
Gary's mark was average in the course, yet he performed very low on the work habit rating. This would seem to suggest that there is no relationship between his average score and his work habit rating.

In the case of Jerry, his average mark is fairly high while his work habit rating is low. This shows that you cannot always be sure about the relationship as predicted by the correlation coefficient.



4

4. Find the line of best fit for the following scatterplot using the median fit method.



In this case there are fifteen points. The points are divided into three strips so that there are five points in each strip. The median is found for each strip.

Strip 1 has the points (10, 20), (10, 30), (20, 15), (25, 20), and (30, 10).

The x-coordinates are 10, 10, 20, 25, and 30. The median is 20.

The y-coordinates are 10, 15, 20, 20, and 30. The median is 20.

The median point for Strip 1 is (20, 20).

Strip 2 has the points (40, 25), (40, 40), (50, 30), (50, 35), and (60, 45).

The x-coordinates are 40, 40, 50, 50, and 60. The median is 50.

The y-coordinates are 25, 30, 35, 40, and 45. The median is 35.

The median point for Strip 2 is (50, 35).

Strip 3 has the points (70, 60), (75, 70), (75, 80), (80, 70), and (90, 80).

The x-coordinates are 70, 75, 75, 80, and 90. The median is 75.

The y-coordinates are 60, 70, 70, 80, and 80. The median is 70.

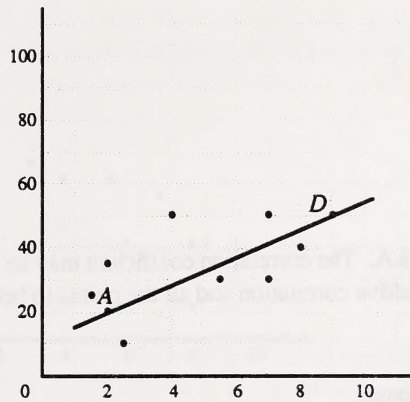
The median point for Strip 3 is (75, 70).

Plot these three points (20, 20), (50, 35), and (75, 70) on the scatterplot and mark them with an X.

Then place a ruler on the median points of the two outside strips. Now slide the ruler down parallel to the line between the first two points. Slide it down one third of the distance to the median point in the middle strip. Draw a straight line here. This is the line of best fit.

5

5. Determine the equation for the line of best fit  $AD$  on the following scatterplot.



Use points  $A(2, 20)$  and  $D(9, 50)$  to find the slope of the line of best fit.

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

$$m = \frac{50 - 20}{9 - 2}$$

$$m = \frac{30}{7}$$

Substitute the slope and the coordinates of point  $A$  into the point slope form of the linear equation.

$$y - y_1 = m(x - x_1)$$

$$y - 20 = \frac{30}{7}(x - 2)$$

$$7y - 140 = 30x - 60$$

$$0 = 30x - 7y + 80$$

The equation of the line is  $30x - 7y + 80 = 0$ .

- 3 6. Circle the correct response.  
Which of the following is the range of the correlation coefficient?

A.  $-1 \leq r \leq 1$

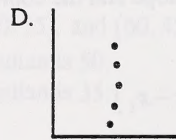
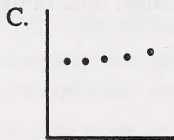
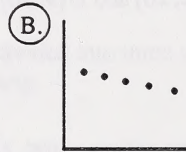
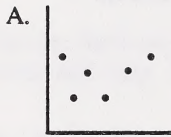
B.  $0 \leq r \leq 1$

C.  $-1 < r \leq 1$

D.  $-1 \leq r \leq 0$

The correct response is A. The correlation coefficient may be  $-1$  for a strong negative correlation and  $+1$  for a strong positive correlation and all the points in between.

- 3 7. Circle the correct response.  
Which of the following graphs is an example of a strong negative correlation?

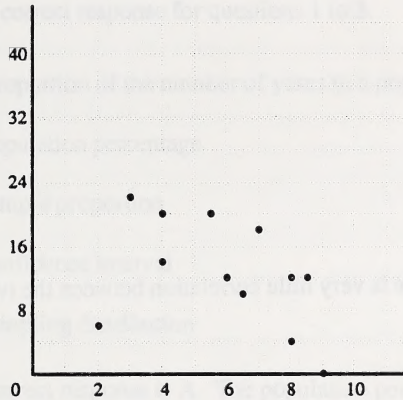


The correct response is B.



5

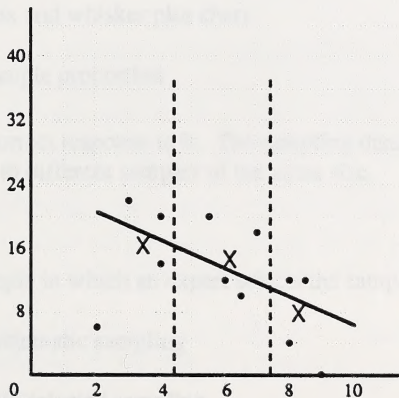
8. Use the following scatterplot to make a table of values.



x	2	3	4	4	5.5	6	6.5	7	8	8	8.5	9
y	6	22	14	20	20	12	10	18	4	12	12	0

4

9. Draw the line of best fit for the following scatterplot.



The median point for Strip 1 is (3.5, 17).

The median point for Strip 2 is (6.25, 15).

The median point for Strip 3 is (8.25, 8).

Place your ruler on the median points in the two outside strips; then slide your ruler up to one-third of the distance to the median point in Strip 2. Draw the line. This is the line of best fit.

- ③ 10. Circle the correct response.  
Which of the following is the weakest correlation coefficient?

- A. + 0.6
- B. - 0.55
- C. + 0.12
- ④ D. + 0.04

The correct response is D. There is very little correlation between the two sets of data when the correlation coefficient is 0.04.

- ③ 11. Circle the correct response.  
Which of the following indicates how well the line of best fit demonstrates the situation?

- A. standard deviation
- ④ B. correlation coefficient
- C. confidence coefficient
- D. sample of the population

The correct response is B.

## Topic 1

\_\_\_\_\_ marks

## Topic 2: Confidence

Circle the correct response for questions 1 to 3.

②

1. The proportion of the number of yeses in a population is called the

- ☒ (A.) population percentage
- B. sample proportion
- C. confidence interval
- D. sampling distribution

The correct response is A. The population percentage is the percentage (or proportion) of objects in the population that are yeses.

②

2. The table that is used to make a box and whisker plot is called the

- A. confidence interval
- ☒ (B.) sampling distribution
- C. box and whisker plot chart
- D. sample proportion

The correct response is B. The sampling distribution is a table that keeps track of the number of yeses in different samples of the same size.

②

3. A sample in which an expert selects the sample based on knowledge of the population is called

- A. systematic sampling
- B. self-selected sampling
- C. clustered sampling
- ☒ (D.) judgement sampling

The correct response is D.



## 4. Explain the following.

4

## a. correlation

The correlation between two or more variables is a measure of the relationship between the variables. If two variables are correlated, they generally have a lot in common. Mathematics and science, for example, should have a strong correlation. On a grid, a positive correlation has its graph rising to the right. A negative correlation has its graph falling to the right.

A measure of how well two variables are related is given by the correlation coefficient. The range of this coefficient is between  $-1$  and  $+1$  or  $-1 \leq r \leq 1$ .

It should be noted that  $-1$  represents a very strong negative correlation and  $+1$  represents a very strong positive correlation.

4

## b. bivariate

A set of data that contains two variables is said to be bivariate. For example, if you compare the wingspan of an aircraft with the cost of the aircraft, you would be dealing with two variables that are not closely related yet have a correlation. The correlation may be negative or positive, but the correlation does exist.

A study of this bivariate data would produce a scatterplot graph, and a line of best fit would indicate the correlation.

4

## c. line of best fit

Sometimes the points on a scatterplot graph appear to be clustered about a line. The straight line that can be drawn joining the greatest possible number of points on the scatterplot is called the line of best fit.

The slant of the distribution can be determined by this line. The line of best fit is usually sketched so that an equal number of points lie on either side of the line. If the points are close to the line, you have strong correlation between the variables. A strong correlation may be either negative or positive.

4

## d. 95% box and whisker plots

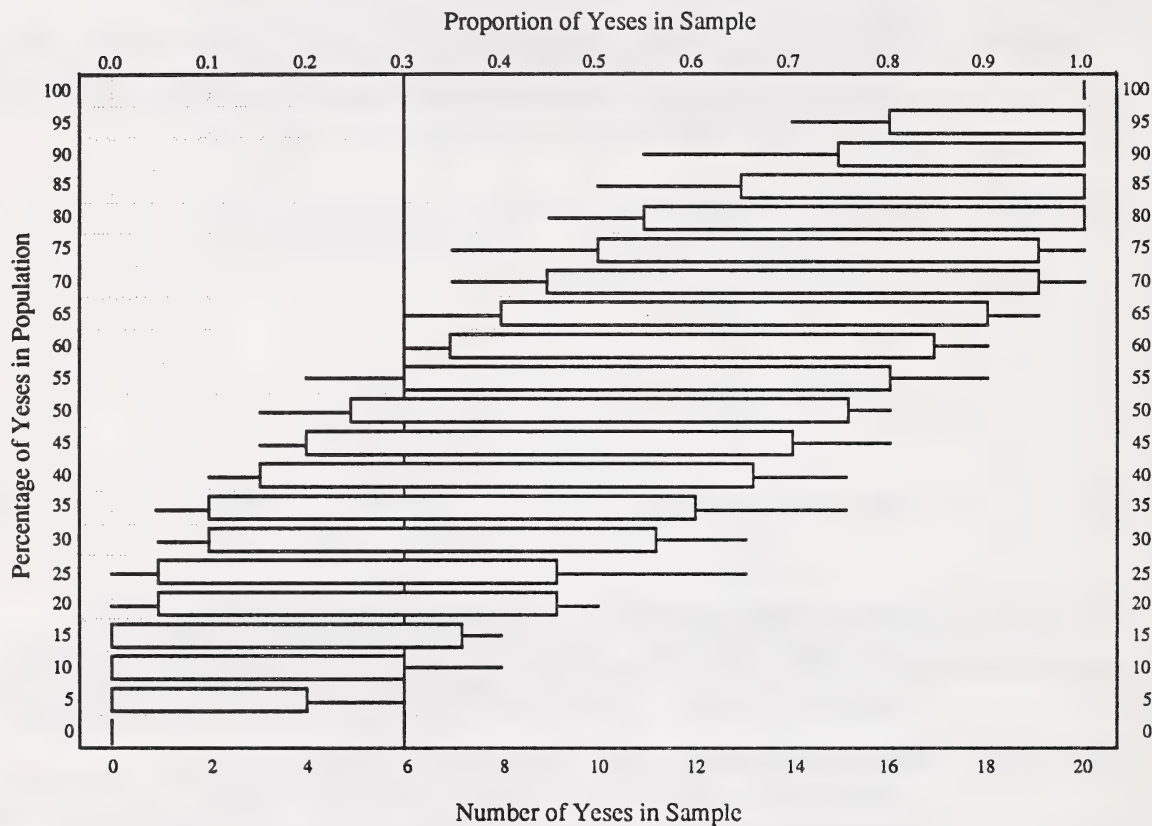
Surveys and polls are often designed to collect yes or no responses to a question from samples within a population. The number of positive responses in each of the samples is used to calculate the population percentage, which is the percentage of the entire population that would respond positively. This information is then organized into a box and whisker plot.

The 95% box and whisker plot is designed so that when further samples of the same size are taken from the population, 95 out of 100 samples will have the number of positive responses as indicated by the box in the plot. Five samples out of 100 will be outside of the box.

A box and whisker plot is one way of showing the amount of confidence that you have in predicting the outcome of a sample of a certain size.

5. A survey of parents was conducted to find out if they thought they were failures at parenting. The box and whisker plot charts that follow show the results of the survey. Study this chart and answer the questions that follow.

**90% Box and Whisker Plots from Samples of Size 20**





②

- a. How many parents answered yes?

Six respondents said yes, they thought they were failures at parenting.

②

- b. How many parents were polled?

The survey interviewed a random sample of twenty parents.

②

- c. What is the sample proportion?

The sample proportion, according to the chart, is 0.3.

②

- d. What percentage of box and whisker plots was used?

The surveyor used a 90% box and whisker plot from a sample size of 20.

②

- e. What is the confidence interval?

According to the chart, 10% to 55% is a 90% confidence interval. In other words, between 10% and 55% would answer yes to the question 90% of the time.

②

- f. What are the lower and upper confidence limits?

The lower confidence limit is 10% and the upper confidence limit is 55%.

④

6. Construct a sampling distribution using the following data.

Sample Number	Responses	Number of Yeses
1	NNYNNNNYNNNYNNNNNNNN	4
2	YNNYNNNNYNNNNNNNNNNY	6
3	YYYNNNNNNYNNNNYNNYNY	8
4	NNNYNNNNYNNNNNNNNYNN	4
5	YYYNNNNNNYNNNNNNNNYYY	8
6	NNNNNNNNYNNNNYNNNNYNN	5
7	YYYNNNYNNYYYNNNNNNYNN	10
8	NNNYNNNNNNYNNNNNNYNN	6
9	NYNNNNYNNNNYNNNNNNYNN	7
10	NNYNNNNYNNNNYNNNNYNNY	7
11	NYNNNNNNYNNNNNNNNNNYNN	5
12	YYYNNNNNNNNNNYNNYYYNN	8
13	NNNYYYYNNNNYNNNNNNNY	9
14	YNNNNYNNNNYNNNNYNNNY	7
15	NNNNNNYNNYYYNNYYYNNYNN	9
16	NNNYNNNNYNNNNYNNNNYNN	6
17	NNYNNNNYNNNNYNNNNNNYY	6
18	NNNNNNNNYNNNNYYYNNNNYY	7
19	NNNNNNYNNNNYYYNNNNYNN	7
20	YYYNNNNNNYNNNNNNNNNN	5

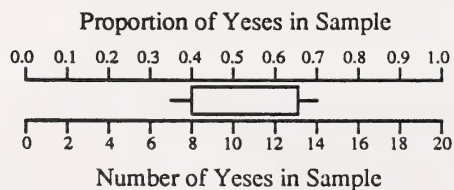
Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	0	0.00
1	0.05	0	0.00
2	0.10	0	0.00
3	0.15	0	0.00
4	0.20	2	0.10
5	0.25	3	0.15
6	0.30	4	0.20
7	0.35	5	0.25
8	0.40	3	0.15
9	0.45	2	0.10
10	0.50	1	0.05
11	0.55	0	0.00
12	0.60	0	0.00
13	0.65	0	0.00
14	0.70	0	0.00
15	0.75	0	0.00
16	0.80	0	0.00
17	0.85	0	0.00
18	0.90	0	0.00
19	0.95	0	0.00
20	1.00	0	0.00
Total		20	1.00



4

7. Use the following sample distribution to make a 90% box and whisker plot.

Number of Yeses	Sample Proportion	Frequency	Proportion of All Trials
0	0.00	0	0.00
1	0.05	0	0.00
2	0.10	0	0.00
3	0.15	0	0.00
4	0.20	0	0.00
5	0.25	0	0.00
6	0.30	0	0.00
7	0.35	1	0.05
8	0.40	1	0.05
9	0.45	4	0.20
10	0.50	5	0.25
11	0.55	3	0.15
12	0.60	3	0.15
13	0.65	2	0.10
14	0.70	1	0.05
15	0.75	0	0.00
16	0.80	0	0.00
17	0.85	0	0.00
18	0.90	0	0.00
19	0.95	0	0.00
20	1.00	0	0.00
Total		20	1.00



8. Fill in the following blanks.

②

a. A **random sample** is selected when each member of the population has an equal chance of being selected and the members of the sample are chosen independently.

②

b. A sample is said to be **biased** if a selection method tends to overrepresent or underrepresent some part of the population.

②

c. Judgement and self-selected samples are both types of **convenience** sampling.

②

d. A **stratified random** sample occurs when the population is divided into different strata, and then a random sample is taken from each strata.

## Topic 2

\_\_\_\_\_ marks







NLC - BNC



3 3286 10840158 5





